Environment Canada Environnement Canada

Regional Frequency Analysis Guidelines

- The Flood Frequency Case -

by

Paul J. Pilon

Environment Canada
Inland Waters Directorate
Water Resources Branch
Ottawa, Ontario
K1A 0H3

and

W.O. Thomas, Jr.

United States Department of the Interior
Geological Survey
Office of Surface Water
Reston, Virginia 22092

February 1990

Canada

TABLE OF CONTENTS

	Page
Summary	ii
1. Introduction	1
2. Single-Site Frequency Analysis	2
2.1 Flood Frequency at a Gauged Location	2
3. Regionalization Procedures	5
3.1 Index-flood Procedure	6
3.2 Ordinary Least Squares Regression Procedure	7
3.3 Generalized Least Squares Regression Procedure	8
4. Summary of Regionalization Procedures	9
5. References	11

SUMMARY

The Water Resources Branch provides Surface Water data and information to hydrologists, water managers, and other interested users in Canada. Part of the information developed relates to tools and procedures for interpreting streamflow data in different basins and regions.

This document provides a guideline for performing a regional flood analysis. It is not intended as an exhaustive description of the approaches which may be taken, but rather as a guide to the procedures that are commonly applied within the Water Resources Branch and similar agencies. Step-by-step instructions give the sequence by which studies are normally conducted. References are made to appropriate sources for further study.

1. INTRODUCTION

Streamflow and water level data from approximately 3300 gauging stations are published by the Water Resources Branch of Environment Canada. Data from these gauges provide information vital to the management of our water resources. Traditionally, the data collected at a specific site is used for various purposes such as streamflow forecasting, flood plain management, and the design of hydraulic structures, to name but a few. It is, however, economically impossible to acquire historical records for the tens of thousands of locations where streamflow data may be required. For maximum usefulness, information derived from records must be made applicable to ungauged areas. Regionalization procedures are used to transfer streamflow (annual means, annual flood peaks, n-day lowflows, etc.) characteristics from gauged to ungauged basins. This represents an extension of the gauging network and allows planners and designers to make estimates of the frequencies of streamflow characteristics at ungauged sites of interest.

Regionalization procedures usually have as a basis the computation of the frequency of occurrence of the streamflow characteristic for the gauged locations. This paper first presents a method of analyzing streamflow data to obtain information about the magnitude and frequency of flood discharges. It then briefly describes three regionalization procedures which are used within the Water Resources Branch. They are: (1) the index-flood procedure; (2) the ordinary least squares (OLS)

regression procedure; and (3) the generalized least squares (GLS) regression procedure. The paper then gives a step-by-step description of the regionalization process.

2. SINGLE-SITE FREQUENCY ANALYSIS

The Consolidated Frequency Analysis (CFA) Package (8) and the Lov Flow Frequency Analysis (LFA) Package (9) reflect the latest developments used in the Water Resources Branch. The methods contained in those packages will be revised periodically, as would be expected in such a rapidly developing subject.

The discourse which follows is directed to flood analysis rather than annual mean or n-day lowflow analyses. The philosophy of regionalizing the three types of streamflow characteristics is very similar, but emphasis is placed on floods in order to simplify this document. Other streamflow characteristics will be addressed in later documents.

2.1 Flood Frequency at a Gauged Location

The first step in a regionalization study is to select the hydrometric sites for analyses. Sites should have at least 10 years of record, although records as short as five years have been used (particularly when the regionalization procedure is generalized least squares regression). Sites with storage or other artificial factors, which would tend to modify flood discharges significantly, should be identified and kept to a minimum. Where discharge is subject to

noticeable control by humans, the sites should be avoided. The records used in the study should be representative of natural conditions in the region.

The artificial controlling of discharge should be evident in a review of the station's history. Records prior to the alteration of the flow regime may be acceptable for regionalization purposes, provided they represent natural conditions. A review of the station's history should also include a thorough review of the stage-discharge rating. This should be performed by persons experienced in field and office stream-gauging practice. The desired goal is to obtain a consistent set of peak discharges that are as accurate as possible. A common source of inconsistency arises when changes in rating curves are made on the basis of recent discharge measurements, while past records are not reviewed or revised.

The nonparametric (test for trends, independence, etc.) and visual data screening (plots of discharge versus time, etc.) tools of the CFA package are intended to complement the review of the station's history. A thorough knowledge of basin activities that influence discharge estimation (the stage-discharge relationship, the climatological factors governing the generation of each flood peak such as snowmelt and/or rainfall, significant changes in land-use, construction of retention ponds, etc.) are of extreme importance to the assurance of the consistency of a station's data.

Flood data are usually in the form of maximum daily discharges or maximum instantaneous discharges. Depending on the information that is being transferred, the study can be focused on one or the other. If the study is to emphasize the maximum instantaneous discharges, some prefer to regionalize the maximum daily discharges and use a formula to scale or transfer the results to instantaneous values. This approach should be adopted with extreme caution as there may not always be a physical relationship between the maximum daily and instantaneous discharges for all basins in a region. In addition, the use of the transfer model adds additional error to the regionalization model. This may negate the additional years of data that are available when basing an analysis on maximum daily discharges. One method to compensate for this loss is by the use of historic information. It may be estimated at the site from an analysis of the maximum instantaneous and daily discharges.

The CFA package performs a frequency analysis on the annual flood peaks and provides a table of design floods and a graph of the selected cumulative distribution function. The CFA documentation (8) can be consulted for more information on single-site analysis.

Single-site frequency analysis is performed on all sites with a suitable station history and a sufficient record length. These sites will have also successfully passed data screening procedures. Results will be given in numerous tables and graphs of flood magnitude versus recurrence internal for each station.

3. REGIONALIZATION PROCEDURES

The information generated from single-site analysis is not applicable directly to ungauged areas. In most cases, the flood records are short and have correspondingly high sampling errors. Often, the records represent different periods of time.

It is assumed that the peaks at any one station form a random sample. As such, they may be far different in character, particularly in a short record, from those of nearby stations otherwise of like physical characteristics. It is also recognized that the sites in a region may somehow be linked from year to year by a certain degree via flood generation mechanism. In other words, the sites do not contain data which are truly independent. This lack of independence should be considered in regionalization studies.

There are two aspects to the analysis of flood frequency in a region: (1) time sampling and (2) spatial or geographic sampling. For the first, we require a sample of a large number of years, not necessarily at one station. For the second, we need a large number of stations to sample a wide diversity of catchment characteristics.

The three regionalization procedures mentioned in the introduction differ in the way they treat the above mentioned two aspects. Their differences are drawn out in the description of the procedure that follows.

3.1 Index-Flood Procedure

This procedure represents the combination of two major parts. The first deals with the development of a dimensionless frequency curve. It represents the ratio of flood discharges at selected recurrence intervals to an index flood, such as the mean annual flood or the flood corresponding to a specified return period. The second part represents the development of a relationship between watershed characteristics (climatic, physiographic) and the index flood. This relationship allows the index to be estimated at any point in the region. The combination of the two parts yields a frequency curve at any location.

The dimensionless frequency curve results from: (1) a frequency analysis of each station based on a prescribed distribution; (2) the formulation of homogeneous regions, based on a homogeneity test of the slopes of the frequency curves; and (3) the derivation of the regional dimensionless frequency curve, based on median flood ratios for certain recurrence intervals for each station in the region. The homogeneity test uses the ratio of the 10-year flood to the mean annual flood to determine whether the differences in slopes of frequency curves for all stations in a given region are greater than those attributed to chance.

The above procedure assumes that the slopes of the frequency curves within a homogeneous region do not vary with watershed characteristics. This assumption must be tested; otherwise, the model may inadequately represent frequency curves within the region.

The index flood at ungauged locations is estimated based on a relationship between the index at gauging stations and their basins' characteristics. Some characteristics which are often used include drainage basin size, mean basin elevation, mean annual precipitation, area of lake and swamps, and hydrologic radius (basin area divided by basin perimeter). This relationship should, as well, be regionalized so that it may be applied in a reliable fashion. The homogeneous regions of the dimensionless frequency curve and the index flood need not coincide.

The index-flood procedure was popularized by the U.S.G.S. (3). Their procedure assumed the parent distribution of floods to be the Gumbel I or extreme value type I distribution. Condie (1) extended this procedure by deriving an homogeneity test for the three parameter lognormal distribution. Condie and Harvey (2) document the adaptation of the three parameter lognormal distribution to the index flood approach. They also document a computer program entitled RFA3LN to assist in the application of their index procedures and provide an example.

3.2 Ordinary Least Square Regression Procedure

Flood discharges at selected recurrence intervals (T-year floods) are related to watershed characteristics using multiple regression techniques. Single-site frequency analysis is used to determine the T-year floods. The "best" estimate for a particular return period may be selected using judgement. Alternatively, the parent distribution

may be assumed **a priori**. The regression equations that relate the T-year floods to basin characteristics are based on the ordinary least squares fitting method. Equal weights are given to all stations in the analysis regardless of the record length, the accuracy of the quantile estimate, and the possible replication of information to various degrees among sites. Several texts (4-6) describe the ordinary least square fitting method. A Water Resources Branch publication (7) can also be consulted for more information on the method.

Homogeneous regions are generally determined from a visual assessment of the areal distribution of regression residuals. The definition of the region is dependent on both the selected recurrence interval and the chosen model. Hence, each equation could have its own unique region.

3.3 Generalized Least Squares Regression Procedure

This procedure is similar to that outlined for the ordinary least squares (OLS) regression procedure. The difference is the OLS regression procedure does not account for variable errors in flood characteristics that are, in part, due to the unequal record lengths at gauging stations. The generalized least squares (GLS) regression procedure accounts for both the unequal reliability and the correlation of flood characteristics among sites. In addition, the GLS procedure provides more accurate estimates of the model's coefficients and their corresponding accuracy, as well as better estimates of the model's error than does the OLS regression procedure (10, 12, 13).

The current GLS procedure assumes that the floods are drawn from a log Pearson Type III distribution (11). It also assumes that the form of the regression model is log-linear. That is, the dependent variable is always transformed logarithmically in the regional model.

The definition of homogeneous regions resulting from a GLS model is theoretically more complex than that described for the OLS. It is a result of residuals being associated somewhat with site reliability. Due to this complexity, current procedures suggest establishing both the region and the form of the regression model using OLS. Once these have been established, GLS procedures are applied.

4. SUMMARY OF REGIONALIZATION PROCEDURES

The step-by-step procedure for determining regional flood frequencies is as follows:

(1) Obtain flood data for all gauges in and around the region of interest. The record should be a minimum of five years.

(2) Review each station's history and make careful study of the stage-discharge relationship. The data should be: as error-free as possible; not significantly influenced by human activity; free of trends and jumps; and homogeneous with respect to both time and flood mechanism. The screening tests within CFA can be used to help assess a portion of the above.

(3) Research, assess, and add, where possible, historical information for each station to the data base. A Water Resources Branch publication (8) can be consulted for more information regarding the definition of historical information and how it is used in frequency analysis.

(4) Select the regionalization procedure.

(5) Quantify the characteristics for each basin that influence the flood discharges at the recurrence intervals of the study. OLS/GLS procedures may have several recurrence intervals of interest, while the index-flood procedure is concerned with only the recurrence interval associated with the index.

(6) Perform a single-site frequency analysis on all sites retained after step (2). In Water Resources Branch, the index-flood procedure is based on the three parameter lognormal distribution (Program RFA3LN), while the GLS procedure is based on the log Pearson Type III distribution (LP3NET). Otherwise, the CFA Package is used to obtain the best estimate of the T-year flood for the OLS regression procedure.

(7) If an index-flood approach is being performed, then the test for homogeneity must be peformed regarding the slope of the index curve. Also, verify that the slope does not vary as a function of basin characteristics. If it does, then a more complex modelling scheme is justified.

(8) Determine the OLS regression model to relate the index flood to basin characteristics (4-7). This includes the postulation of the form of the model, the selection of the dependent and independent variables, analysis of the pattern of the residuals, and the formulation of the homogeneous region for application of the model. This procedure is repeated for each additional recurrence interval when OLS/GLS regression procedures are used to estimate the flood frequency regime. Note that current GLS procedures assume the dependent variable to be always logarithmically transformed.

5. REFERENCES

1. Condie, R. The Three Parameter Lognormal Distribution Applied to Regional Flood Frequency Analysis by the Index Flood Method, Environment Canada, Technical Workshop Series No. 2, 1979.

2. Condie, R. and Harvey K.D. Regional Dimensionless Flood Frequency Curves using the Three-Parameter Lognormal Distribution, Environment Canada, Water Resources Branch, March 1985.

3. Dalyrmple, T. Flood Frequency Analysis, USGS Water-Supply Paper 1543-A, 1960.

4. Draper, H. and Smith, H. Applied Regression Analysis, Second Edition, John Wiley & Sons Inc., 1981.

5. Green, P.E. Analyzing Multivariate Data, the Dryden Press, 1978.

6. Neter, J., Wasserman, W., and Kutner, M.H. Applied Linear Regression Models, Richard D. Irwin Inc., 1983.

7. Pilon, P.J. and Condie, R. Multiple Regression in Hydrology, Environment Canada, Water Resources Branch, November 1983.

8. Pilon, P.J., Condie, R., and Harvey, K.D. Consolidated Frequency Analysis Package - User Manual for Version 1 - DEC PRO Series, Environment Canada, Water Resources Branch, July 1985.

9. Pilon, P.J., and Jackson, R.J. Low Flow Frequency Analysis Package - LFA - User Manual for Version 1 - DEC PRO Series, Environment Canada, Water Resources Branch, September 1988.

10. Thomas, W.O. Jr., and Cheng, L.C., Hydrometric Network Analysis in the Oldman River Basin, Alberta using Generalized Least Squares, Proceedings of the 7th Canadian Hydrotechnical Conference of the CSCE, May 27 - June 1, 1985, Saskatoon, Sask.

11. Thomas, W.O. Jr., Cheng, L.C. and Tasker, G.D. Computer Procedures for Hydrologic Regression and Network Analysis using Generalized Least Squares, Environment Canada, Water Resources Branch, May 1985.

12. Stedinger, J.R. and Tasker, G.D. Regional Hydrologic Analysis 1. Ordinary, Weighted, and Generalized Least Squares Compared, Water Resources Research, 21(9), 1421-1432, September, 1985.

13. Stedinger, J.R. and Tasker G.D. Correction to "Regional Hydrologic Analysis ...", Water Resources Research, 22(5). 844, May 1986.